

EXHIBIT A

**UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK**

AUTHORS GUILD, et al.,

Plaintiffs,

v.

OPEN AI INC., et al.,

Defendants.

ECF CASE

No. 1:23-cv-08292-SHS;

No. 1:23-cv-10211-SHS

**PLAINTIFFS' FOURTH SET OF
REQUESTS FOR PRODUCTION
TO OPENAI**

JONATHAN ALTER, et al.,

Plaintiffs,

v.

OPENAI, INC., et al.,

Defendants.

PLAINTIFFS' FOURTH SET OF REQUESTS FOR PRODUCTION TO OPENAI

Pursuant to Rules 26 and 34 of the Federal Rules of Civil Procedure, Plaintiffs request that Defendant OpenAI respond to the following Third Set of Requests for Production of Documents ("Requests"). Responses to these Requests shall be due within thirty (30) days of the date of service or as otherwise mutually agreed by the parties. Plaintiffs are amenable to an electronic production, subject to agreement by the parties.

In accordance with Rule 34(b), OpenAI shall provide written responses to the following Requests and shall produce the requested documents as they are kept in the ordinary and usual course of business or shall organize and label the documents to correspond with the categories in these Requests. In accordance with Rule 26(e), OpenAI shall supplement or correct its responses

or productions as necessary.

I. DEFINITIONS

1. **“Actions”** means the above captioned litigation, *Authors Guild et al. v. Open AI Inc. et al.*, No. 1:23-cv-08292-SHS (S.D.N.Y.) (opened Sept. 19, 2023 and amended December 5, 2023), and *Alter et al. v. OpenAI et al.*, No. 1:23-cv-10211-SHS (S.D.N.Y.) (opened Nov. 21, 2023, and amended Dec. 19, 2023).

2. **“Communication(s)”** means the transmittal of information (in the form of facts, ideas, inquiries or otherwise) by any means, including, but not limited to, telephone calls, emails (whether via company server or personal webmail or similar accounts), faxes, text messages (on work or personal phones), instant messages, Skype, Line, WhatsApp, WeChat, other electronic messages, letters, notes, and voicemails.

3. **“Document(s)”** is defined to be synonymous in meaning and equal in scope to the usage of the term “documents or electronically stored information” in Rule 34(a)(1)(A). For the avoidance of doubt, **Document(s)** includes **Communication(s)**.

4. **“You”, “Your”, and “OpenAI”** means OpenAI, Inc., OpenAI GP, LLC, OpenAI, LLC, OpenAI OPCO LLC, OpenAI Global LLC, OAI Corporation, LLC, OpenAI Holdings, LLC (each, an **“Entity”**), and any of their directors, officers, employees, partners, members, representatives, agents (including attorneys, accountants, consultants, investment advisors or bankers), and any other person acting or purporting to act on their behalf, as well as corporate parents, subsidiaries, affiliates, predecessor entities, successor entities, divisions, departments, groups, acquired entities, related entities, or any other entity acting or purporting to act on their behalf.

5. **“Large Language Model,” “LLM,” “AI Model(s),” “Generative AI system(s),” “model(s),” and “API Product(s)”** have the same meaning as they are used in

YOUR letter to the Register of Copyrights and Director of the U.S. Copyright Office dated October 30, 2023, “Re: Notice of Inquiry and Request for Comment [Docket No. 2023-06]” and include all models listed or described in <https://platform.openai.com/docs/models>.

6. **“Person(s)”** means any individual or entity.

7. **“Fine Tune(d),” “Fine Tuning,” “Pre-Train(ed),” “Pre-Training,” “Train[ed],”** and **“Training,”** have the same meaning as the term is discussed in **Your** website materials and statements. See e.g., <https://platform.openai.com/docs/guides/fine-tuning> (“OpenAI’s text generation models have been pre-trained on a vast amount of text. To use the models effectively, we include instructions and sometimes several examples in a prompt. Using demonstrations to show how to perform a task is often called “few-shot learning.” Fine-tuning improves on few-shot learning by training on many more examples than can fit in the prompt, letting you achieve better results on a wide number of tasks. Once a model has been fine-tuned, you won’t need to provide as many examples in the prompt.”); <https://openai.com/index/language-unsupervised/> (“These results provide a convincing example that pairing supervised learning methods with unsupervised pre-training works very well; this is an idea that many have explored in the past, and we hope our result motivates further research into applying this idea on larger and more diverse datasets.”); and <https://openai.com/index/how-should-ai-systems-behave/> (“The two main steps involved in building ChatGPT work as follows . . . First, we “pre-train” models by having them predict what comes next in a big dataset . . . Then, we “fine-tune” these models on a more narrow dataset that we carefully generate with human reviewers who follow guidelines that we provide them.”); <https://openai.com/index/gpt-4-research/> (“Interestingly, the base pre-trained model is highly calibrated . . . Note that the model’s capabilities seem to come primarily from the pre-training process . . . ”); <https://openai.com/index/text-and-code-embeddings-by-contrastive-pre-training/>

(“In this work, we show that contrastive pre-training on unsupervised data at scale leads to high quality vector representations of text and code.”); <https://openai.com/index/vpt/> (“We trained a neural network to play Minecraft by Video PreTraining (VPT) on a massive unlabeled video dataset of human Minecraft play, while using only a small amount of labeled contractor data. With fine-tuning, our model can learn to craft diamond tools . . .”).

8. **“ChatGPT”** means all consumer-facing versions of the chatbot application OpenAI released in November 2022 and any **LLM** underlying those consumer-facing applications.

9. **“Concern”** means be the subject of, make reference to, comment on, discuss, describe, identify, or contain text or images about the stated topic.

10. **“Including”** means including but not limited to.

11. **“Investors”** means any person or entity that financially invested in OpenAI in the operations and/or entitie(s) controlled and/or overseen by one or more of the OpenAI defendant entities.

12. **“Microsoft”** means Microsoft Corporation and any of its directors, officers, employees, partners, members, representatives, agents (INCLUDING attorneys, accountants, consultants, investment advisors or bankers), and any other person acting or purporting to act on their behalf, as well as corporate parents, subsidiaries, affiliates, predecessor entities, successor entities, divisions, departments, groups, acquired entities, related entities, or any other entity acting or purporting to act on its behalf.

13. **“Relate to”** means refer to, respond to, describe, evidence, or constitute, in whole or in part.

14. **“Publishers”** means a person or entity whose business is publishing materials, including, but not limited to, books, periodicals, magazines, and newspapers.

II. RELEVANT TIME PERIOD

The relevant time period is January 1, 2015 through the present (“Relevant Time Period”), unless otherwise specifically indicated, and shall include all **Documents** and any other information relating to such period, even though prepared or published outside of the Relevant Time Period. If a **Document** prepared before the Relevant Time Period is necessary for a correct or complete understanding of any **Document** covered by any of these Requests, please provide the earlier **Document** as well. If any **Document** is undated and the date of its preparation cannot be determined, please produce the **Document** if it is otherwise responsive to any Request.

III. INSTRUCTIONS

1. The production by one person, party, or entity of a **Document** does not relieve another person, party, or entity from the obligation to produce his, her, or its own copy of that **Document**.

2. Produce **Documents** not otherwise responsive to these Requests if such **Documents** **Relate to** the **Documents** that are called for by these Requests, or if such **Documents** are attached to **Documents** called for by these Requests.

3. Produce each **Document** requested herein in its entirety and without deletion, excisions, redaction, or other modification regardless of whether **You** consider the entire document to be relevant or responsive.

4. If any **Document** is known to have existed but no longer exists, has been destroyed, or is otherwise unavailable, identify the **Document**, the reason for its loss, destruction, or unavailability, the name of each person known or reasonably believed by **Microsoft** to have had possession, custody, or control of the original and any copy thereof (if applicable), and a description of the disposition of each copy of the **Document**.

5. If no **Documents** responsive to a particular Request exist, state that no responsive **Documents** exist.

6. If You assert that any of the **Documents** and things requested are protected from discovery by attorney-client privilege, the attorney work product doctrine, or any other evidentiary privilege, specify for each **Document** (1) the grounds asserted as the reason for non-production; (2) the date the **Document** was prepared; (3) the identity of the attorney(s) who drafted or received the **Document(s)** (if attorney-client privilege or attorney work product protection is claimed); (4) the identity of the parties who prepared or received the **Document**; and (5) the nature of the **Document**.

7. Construe the conjunctions “and” and “or” non-restrictively or non-exclusively if doing so would bring within the scope of these Requests **Documents** that might otherwise be construed to be outside of their scope.

8. Construe the use of the singular to include the plural; the use of the masculine, feminine, or neuter gender to include the others; and the use of one form of the verb to include the others if doing so would bring within the scope of these Requests **Documents** that might otherwise be construed to be outside of their scope.

Plaintiffs, by and through their undersigned attorneys, request that Defendants provide a response to these Requests within thirty (30) days of the date of service hereof as provided by Federal Rule of Civil Procedure 33.

REQUESTS FOR PRODUCTION

REQUEST FOR PRODUCTION NO. 71:

Documents sufficient to show all of **Your** retention policies during the Relevant Time Period and any changes thereto.

REQUEST FOR PRODUCTION NO. 72:

All **Documents Concerning or Relating to** the Data Working Group, as referenced in the Second Amended and Restated Joint Development and Collaboration Agreement, MSFT_AICPY_000000575 (including all predecessor or successor groups), including but not limited to all **Communications** among members of, meeting minutes of, and data policies compiled by the Data Working Group.

REQUEST FOR PRODUCTION NO. 73:

All **Documents Concerning or Relating to** the Governing Board, as referenced in the Second Amended and Restated Joint Development and Collaboration Agreement, MSFT_AICPY_000000575 (including all predecessor or successor groups), including but not limited to all Communications among members of and meeting minutes of the Governing Board.

REQUEST FOR PRODUCTION NO. 74:

Documents sufficient to show the first time at which **You** sought to hire and did hire employees whose job function was primarily the commercialization, monetization or productization of **Your LLMs**.

REQUEST FOR PRODUCTION NO. 75:

Documents sufficient to show each individual employed by OpenAI, Inc., OpenAI GP, LLC, OpenAI, LLC, OpenAI Global LLC, OAI Corporation, LLC, OpenAI Holdings, LLC, the job function of each employee, the supervisory relationship of each employee, and any change in each employee's job function, supervisory relationships, or employment status over time.

REQUEST FOR PRODUCTION NO. 76:

Documents sufficient to show the identity of each board member of OpenAI, Inc., OpenAI GP, LLC, OpenAI, LLC, OpenAI OPCO LLC, OpenAI Global LLC, OAI Corporation, LLC,

OpenAI Holdings, LLC, including any change in that individual's status as a board member over time.

REQUEST FOR PRODUCTION NO. 77:

Documents sufficient to show each physical book that **You** scanned for use as training data in **Your LLMs**, including but not limited to the price paid for each such book, as applicable.

REQUEST FOR PRODUCTION NO. 78:

Documents sufficient to show each digital book that **You** purchased and used as training data in **Your LLMs**, including but not limited to the price paid for each such book.

REQUEST FOR PRODUCTION NO. 79:

Documents sufficient to show the number of reproductions made of training data in the course of training **Your LLMs**.

REQUEST FOR PRODUCTION NO. 80:

Documents sufficient to show any exclusive or non-exclusive rights granted by **You** to Microsoft to commercialize, monetize, or productize **Your LLMs**, training data, or intellectual property.

REQUEST FOR PRODUCTION NO. 81:

All minutes of meetings of the board of directors of each **Entity**, and all **Documents** provided to the boards of directors, that **Concern** or **Relate to Your** training data; **Your** intention to commercialize, monetize, or productize **Your LLMs**; or projections regarding revenues or profits for particular products or for **You** as a whole.

REQUEST FOR PRODUCTION NO. 82:

All instruments effectuating Microsoft's investments into and acquisition of **You**.

REQUEST FOR PRODUCTION NO. 83:

Documents Concerning or Relating to the justifications for the formation of each **Entity**.

REQUEST FOR PRODUCTION NO. 84:

Documents sufficient show the date each **Entity** was formed.

REQUEST FOR PRODUCTION NO. 85:

Documents sufficient to show the date that assets were first contributed or otherwise transferred to each newly formed **Entity**.

REQUEST FOR PRODUCTION NO. 86:

All agreements between or among **Entities**.

REQUEST FOR PRODUCTION NO. 87:

All **Documents Concerning or Relating to** the contractual agreements and ownership structures between or among **Entities**.

REQUEST FOR PRODUCTION NO. 88:

All **Documents**—including contracts and agreements—**Concerning or Relating to** Microsoft's control over **You**.

REQUEST FOR PRODUCTION NO. 89:

All **Documents**—including contracts and agreements—**Concerning or Relating to** control by over **You**.

REQUEST FOR PRODUCTION NO. 90:

All instruments effectuating or memorializing the contribution, transfer, or assignment of assets or liabilities among **Entities**.

REQUEST FOR PRODUCTION NO. 91:

Documents Concerning or Relating to the flow of cash, losses, and profits between and among **Entities**.

REQUEST FOR PRODUCTION NO. 92:

All **Documents Relating to** operational meetings and information exchanged among **Entities**.

REQUEST FOR PRODUCTION NO. 93:

Financial statements, including balance sheets and statements of income and cash flow, for each **Entity**.

REQUEST FOR PRODUCTION NO. 94:

Each **Entity**'s operating agreements, limited liability company agreements, or similar governing documents, including all amendments.

REQUEST FOR PRODUCTION NO. 95:

All **Documents** relating to your June 22, 2023 response to Senator Thomas Roland Tillis's questions to **You Relating to** content creators.

REQUEST FOR PRODUCTION NO. 96:

Documents sufficient to identify the bases of your "aware[ness] that creators are concerned about being replaced or displaced by AI."

REQUEST FOR PRODUCTION NO. 97:

Documents Relating to your "continue[d]engage[ment]" with writers.

Dated: July 16, 2024

Rohit D. Nath

Justin A. Nelson (*pro hac vice*)
Alejandra C. Salinas (*pro hac vice*)
SUSMAN GODFREY L.L.P.
1000 Louisiana Street, Suite 5100
Houston, TX 77002
Tel.: 713-651-9366
jnelson@susmangodfrey.com
asalinas@susmangodfrey.com

Rohit D. Nath (*pro hac vice*)
SUSMAN GODFREY L.L.P.
1900 Avenue of the Stars, Suite 1400
Los Angeles, California 90067
Tel.: 310-789-3100
rnath@susmangodfrey.com

Charlotte Lepic
J. Craig Smyser
SUSMAN GODFREY L.L.P.
1901 Avenue of the Americas, 32nd Floor
New York, New York 10019
Tel.: 212-336-8330
clepic@susmangodfrey.com
csmyser@susmangodfrey.com

Rachel Geman

Rachel Geman
LIEFF CABRASER HEIMANN & BERNSTEIN,
LLP
250 Hudson Street, 8th Floor
New York, New York 10013-1413
Tel.: 212-355-9500
rgeman@lchb.com

Reilly T. Stoler (*pro hac vice forthcoming*)
LIEFF CABRASER HEIMANN & BERNSTEIN,
LLP
275 Battery Street, 29th Floor
San Francisco, CA 94111-3339
Tel.: 415-956-1000
rstoler@lchb.com
ibenserg@lchb.com

Wesley Dozier (*pro hac vice*)
LIEFF CABRASER HEIMANN & BERNSTEIN,
LLP
222 2nd Avenue, Suite 1640
Nashville, TN 37201
Tel.: 615-313-9000
wdozier@lchb.com

Scott J. Sholder
Scott J. Sholder
CeCe M. Cole
COWAN DEBAETS ABRAHAMS & SHEPPARD
LLP
41 Madison Avenue, 38th Floor
New York, New York 10010
Tel.: 212-974-7474
ssholder@cdas.com
ccole@cdas@com

Attorney for Plaintiffs and the Proposed Class

CERTIFICATE OF SERVICE

I hereby certify that on July 16, 2024, a copy of the foregoing was served via electronic mail to all counsel of record in this matter.

Reetu Sinha
(Signature)